

THE ASOMO FEATURES

As a result of the extensive work carried out by Socialware[©] on dealing manually with the ASOMO problem in the recent past, high levels of experience and understanding of determining the major factors that characterise the customer’s attitude have been gained. These factors are the following: (1) The implication of the author with the other customers in the forum, (2) the position of authority of the customer, and (3) the subjective language used in the text.

These broad factors have been helpful in detecting a list of 14 morphological features which characterise each analysed document. In order to engineer this list of features, each document is preprocessed using an open source morphological analyser [1][2]. Firstly, spelling in the entire corpus is checked. Then, the analyser provides information related to the part-of-the-speech (PoS) tagging [4]. Once the preprocessing task is performed, determining the values of the features is just looking for specific patterns in the corpus. In the following paragraphs, a detailed introduction of each factor is given, as well as a description of the features used in each factor.

	Feature	Description	Example	Translation
1	<i>First Persons</i>	Number of verbs in the first person.	Contraté [...]	I hired [...]
2	<i>Second Persons</i>	Number of verbs in the second person.	Tienes [...]	You have [...]
3	<i>Third Persons</i>	Number of verbs in the third person.	Sabe [...]	He knows [...]
4	<i>Relational Forms</i>	Number of phatic expressions, i.e. expressions whose only function is to perform a social task.	Hola, Gracias de antemano, etc.	Hello, Thanks in advance, etc.
5	<i>Agreement Expressions</i>	Number of expressions that show agreement or disagreement.	Estoy de acuerdo contigo, No tienes razn, etc.	I agree with you, you’re wrong, etc.
6	<i>Request</i>	Number of sentences that express certain degree of request.	Me gustaría saber [...], Alguien podría [...]	I’d like to know [...], Would appreciate if someone could [...]

TABLE 1. Subset of features related to the implication of the author with other customers.

The implication of the author. This factor covers the features that are related with the interaction between the author and the other customers in the forum. It consists of six

different features that are described in Table 1. Per each feature, we show its description and an example (with its translation into English) of the type of pattern that matches with the feature.

The position of authority of the opinion holder is mainly characterised by the purpose of the written post and it is related to the potential influence in the readers of the forum. The author could express advice, disapproval with a specific product, predictions, etc. Table 2 shows the six features that are part of this major factor.

	Feature	Description	Example	Translation
7	<i>Imperatives</i>	Number of imperative verbs in the second person.	No compres.	Do not buy
8	<i>Exhorts and Advices</i>	Number of exhort verbs, e.g. recommend, advice, prevent, etc.	Te recomiendo [...]	I recommend that you [...]
9	<i>Sufficiency Expressions</i>	Number of expressions used to corroborate other sentences of the text.	Por supuesto, Naturalmente, etc.	Of course, Naturally, etc.
10	<i>Prediction Verbs</i>	Number of verbs in future.	Voy a probar, Llamaré, etc.	I'm going to try, I'll call, etc.
11	<i>Authority</i>	Number of expressions that denote high degree of authority, usually written in the subjunctive mode	Si fuera tú, [...]	If I were you, [...]
12	<i>Questions</i>	Number of question in the post, both direct and indirect.	¿Qué tal es?, Dime qué te parece, etc.	How is it?, Tell me what do you think of it, etc.

TABLE 2. Subset of features related to the position of authority of the customer.

Subjective language deals with the opinion of the author. In order to determine this factor, we consider solely the adjective detected with the PoS recogniser, as commonly done in the state-of-the-art literature [3]. Then, the adjectives are classified in polarity terms by means of a hand-annotated sentiment-lexicon. As a result of this task, we obtain two features: *Positive Adjectives* and *Negative Adjectives*, which are the number of positive and negative adjectives, respectively, in the text.

The 14 features are normalised to be in the range $[0, 1]$ by dividing them by the maximal observed value.

REFERENCES

- [1] J. Atserias, B. Casas, E. Comelles, M. Gonzalez, L. Padro, and M. Padro. Freeling 1.3: Syntactic and semantic services in an open-source NLP library. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*, pages 48–55, 2006.

- [2] X. Carreras, I. Chao, L. Padro, and M. Padro. An open-source suite of language analyzers. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, volume 10, pages 239–242, 2006.
- [3] V. Hatzivassiloglou and J. M. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th conference on Computational linguistics*, pages 299–305, 2000.
- [4] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'02)*, pages 79–86, 2002.